Structure of the hepatitis E virus-like particle suggests mechanisms for virus assembly and receptor binding

Tom S. Y. Guu[a,1], Zheng Liu[b,1], Qiaozhen Ye[a], Douglas A. Mata[a], Kunpeng Li[c], Changcheng Yin[b], Jingqiang Zhang[c,2], and Yizhi Jane Tao[a,2]

[a]Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77005; [b]Department of Biophysics, Health Science Centre, Peking University, Beijing, China 100191; and [c]State Key Laboratory for Biocontrol, Sun Yat-Sen University, Guangzhou, China 510275

Edited by Michael G. Rossmann, Purdue University, West Lafayette, IN, and approved June 15, 2009 (received for review May 1, 2009)

Hepatitis E virus (HEV), a small, non-enveloped RNA virus in the family *Hepeviridae*, is associated with endemic and epidemic acute viral hepatitis in developing countries. Our 3.5-Å structure of a HEV-like particle (VLP) shows that each capsid protein contains 3 linear domains that form distinct structural elements: S, the continuous capsid; P1, 3-fold protrusions; and P2, 2-fold spikes. The S domain adopts a jelly-roll fold commonly observed in small RNA viruses. The P1 and P2 domains both adopt β-barrel folds. Each domain possesses a potential polysaccharide-binding site that may function in cell-receptor binding. Sugar binding to P1 at the capsid protein interface may lead to capsid disassembly and cell entry. Structural modeling indicates that native T = 3 capsid contains flat dimers, with less curvature than those of T = 1 VLP. Our findings significantly advance the understanding of HEV molecular biology and have application to the development of vaccines and antiviral medications.

capsid | HEV

Viral hepatitis is principally caused by 5 distinct viruses named hepatitis A–E. Despite their similar names, the 5 viruses are unrelated, and they have totally different genome structures with distinct replication mechanisms. Hepatitis E virus (HEV) is responsible for endemic hepatitis as well as sporadic epidemics of acute, enterically transmitted hepatitis in the developing world, including parts of Asia, the Middle East, Africa, and Mexico (1, 2). HEV accounts for more than 50% of acute viral hepatitides in young adults in these regions, with a case fatality of 1–2% in regular patients and up to 20% in pregnant women.

Given the lack of a robust cell culture system, and because HEV is not closely related to any other well-characterized virus, little is known about the molecular biology of HEV or its strategy for replication (1). HEV is a small, non-enveloped virus with a 7.2 kb, positive-sense RNA genome. Its genomic RNA is polyadenylated and contains 3 ORFs. Located near the 5′-end, ORF1 encodes a non-structural polyprotein with multiple functional domains, including those for methyltransferase, protease, helicase, and polymerase. The viral capsid protein (CP) is encoded by ORF2 near the 3′-end. ORF3, which partially overlaps with the other 2 ORFs, codes for an immunogenic protein of unknown function. HEV was originally classified in the *Caliciviridae* family because of its structural similarity to other caliciviruses; however, it is now the sole member of the *Hepeviridae* family. The genomic RNA of HEV exhibits several distinct features compared to the genomic RNA of caliciviruses, including a methylated cap at the 5′-end and an ORF1 with functional domains arranged in a different order (1, 3).

Previous studies of HEV assembly have primarily focused on the overexpression of viral proteins. The ORF2 capsid protein, HEV-CP, contains a total of 660 amino acid residues. At the HEV-CP N terminus is a signal peptide followed by an arginine-rich domain that potentially play a role in viral RNA encapsidation during assembly (3, 4). HEV-CP contains 3 putative N-glycosylation sites (5), but the biological relevance of such potential modifications is unclear (6, 7). The receptor binding site has been mapped to the second half of the polypeptide chain, although the cell receptor for HEV has not yet been identified (8). Expression of the entire ORF2 in insect cells results in proteolytic removal of the first 111 and the last 52 residues (9), yielding a 55-kDa protein capable of self-assembly into virus-like particles (VLPs) (10). The size of the HEV-CP in infectious virions is unknown. HEV-CP is a key antigen that stimulates the host immune response, and 6 antigenic domains have been identified (11). One neutralization site has been mapped to the polypeptide region between amino acids 452 and 617 (12).

The structure of the infectious HEV particle has only been analyzed by immunoelectron microscopy (1). A 22-Å resolution cryo-EM reconstruction has been obtained for the HEV VLP (10). The VLP displays T = 1 symmetry with a diameter of 270 Å, smaller than the native HEV particle, which displays T = 3 symmetry with an estimated diameter of 350–400 Å (10). The surface of the VLP is dominated by 30 dimeric protrusions, and each capsid subunit appears to have 2 domains (10). HEV VLP possesses dominant antigenic activity similar to that of authentic HEV particles, and is therefore a promising candidate for use in vaccine development. A truncated ORF2 polypeptide is currently undergoing clinical trials as a vaccine candidate (13).

Here we report the crystal structure of HEV VLP determined to 3.5-Å resolution. Each HEV-CP contains 3 linear domains, S (118–313), P1 (314–453), and P2 (454–end), the final 2 of which are linked by a long, flexible hinge linker. The S domain forms a continuous capsid shell that is reinforced by 3-fold protrusions formed by P1 and 2-fold spikes formed by P2. It adopts the jelly-roll β-barrel fold that is most closely related to plant T = 3 viruses. P1 and P2 contain compact, 6-stranded β-barrels that resemble the β-barrel domain of phage sialidase and the receptor-binding domain of calicivirus, respectively, both of which are capable of polysaccharide binding. The highly exposed P2 domain likely plays an important role in antigenicity determination and virus neutralization. Structural modeling shows that the assembly of the native T = 3 capsid requires flat capsid protein dimers with less curvatures than those found in the T = 1 VLP, suggesting that additional N terminal sequences may be involved in particle size regulation.

Results and Discussion

Biochemical Characterization of the HEV ORF2 Protein. For crystallization, HEV ORF2$^{112–608}$ was overexpressed in Sf21 insect cells

Author contributions: Q.Y., C.Y., J.Z., and Y.J.T. designed research; T.S.Y.G., Z.L., Q.Y., D.A.M., K.L., J.Z., and Y.J.T. performed research; J.Z. contributed new reagents/analytic tools; T.S.Y.G., Z.L., Q.Y., and Y.J.T. analyzed data; and T.S.Y.G., Q.Y., D.A.M., and Y.J.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID 3HAG).

[1]T.S.Y.G. and Z.L. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: lsszhjq@sysu.edu.cn or ytao@rice.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0904848106/DCSupplemental.

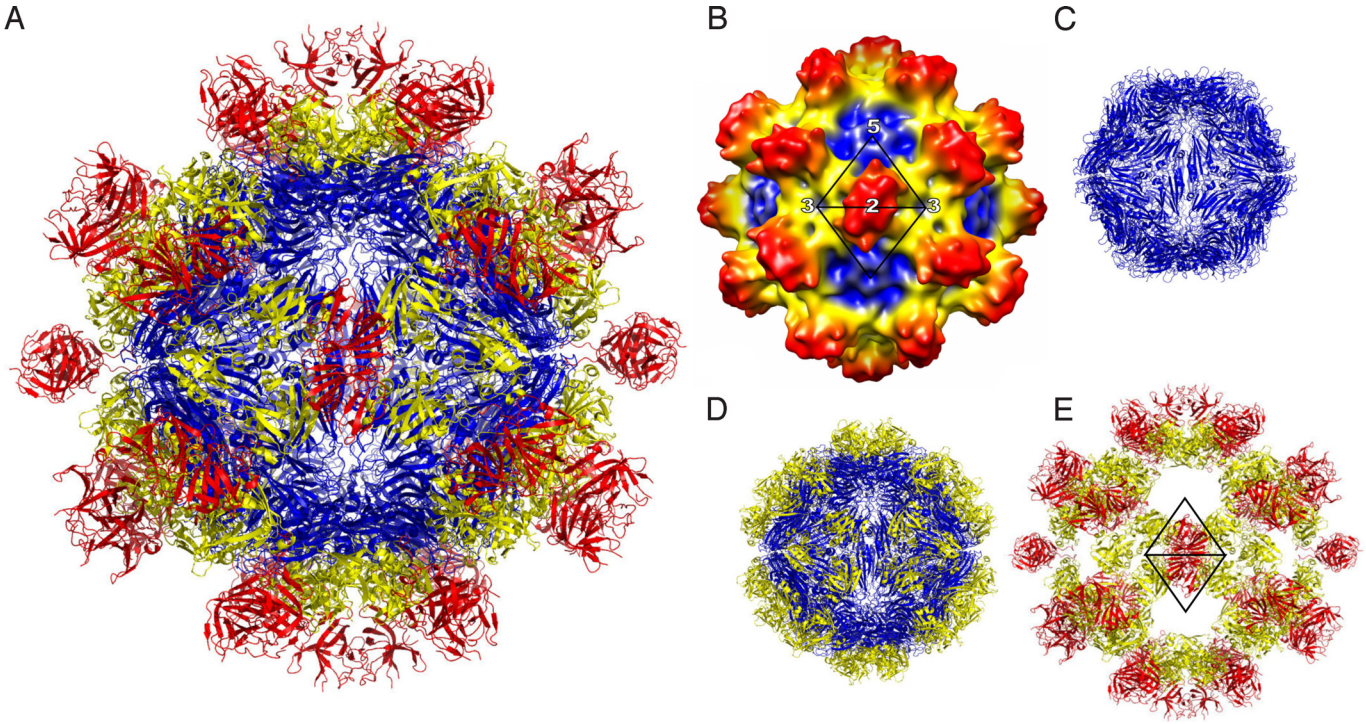

Fig. 1. Structure of the hepatitis E virus-like particle (VLP) (T = 1). (*A*) Crystal structure of HEV VLP. The 3 domains, S, P1, and P2 are colored blue, yellow, and red, respectively. The VLP is positioned in a standard orientation with the 3 2-fold icosahedral symmetry axes aligned along the vertical, horizontal, and viewing directions, respectively. (*B*) Cryo-EM reconstruction at 14 Å resolution. The surface is colored by radial depth cue from blue, yellow, to red. (*C*) HEV VLP with only the S domain. (*D*) VLP with S and P1 domains. (*E*) VLP with P1 and P2 domains.

and purified to near homogeneity by chromatography. According to size-exclusion chromatography, the ORF2$^{112–608}$ protein, hereon referred to as HEV-CP, was purified as dimers with an expected molecular weight of approximately 107 kDa (Fig. S1). Higher molecular weight peaks corresponding to VLPs were not observed, consistent with an earlier report that the overexpression of ORF2$^{112–608}$ in Sf9 cells generates soluble proteins only (10). VLPs were found in the cell media only when Tn5 insect cells were used and infected cells were harvested after prolonged incubation (10). For cryo-electron microscopy, HEV-CP was overexpressed in Tn5 cells, and the VLPs were purified by sucrose gradient ultracentrifugation.

Structure Determination. HEV-CP was crystallized in the space group $P6_3$ with a = 241.1 Å and c = 519.9 Å (Table S1). Unit cell dimensions of this magnitude are unusually large for a 54-kDa protein. Self-rotation function and crystal packing consideration further confirmed that the HEV-CP protein had assembled into a VLP during crystallization, possibly stimulated by the low pH and high ionic strength of the crystallization solution. The structure of the HEV VLP was then determined by phase extension and 20-fold non-crystallographic symmetry (NCS) averaging using a 14-Å cryo-EM reconstruction of the VLP as a phasing model.

Our initial map was of excellent quality with continuous main-chain density and defined side-chain density that allowed the polypeptide chain to be traced without ambiguity (Fig. S2). The side chain features became even more prominent after structure factor sharpening with a negative B factor. Our final model contains 468 out of a total of 497 residues present in the ORF2$^{112–608}$ construct. The missing residues are the first 6, the last 3, and several internal loop regions (148–149, 357–360, 483–488, 574–576, and 589–593). There are 2 cysteines in the ORF2$^{368–606}$ sequence, but neither is involved in intra- or inter-molecular disulfide bonding.

The 3.5-Å structure of the HEV VLP fits well into the 14-Å cryo-EM map (Fig. 1 *A* and *B*), indicating that there is virtually no difference in the structure of VLPs assembled during crystallization and those found in cell media. The crystal structure shows an empty capsid shell with an inner diameter of approximately 125 Å and an outer diameter of approximately 270 Å. The most prominent structural feature is a total of 30 dimeric spikes situated on the icosahedral 2-fold symmetry axes. These dimeric spikes are approximately 30 Å tall. Capsid protrusions are also observed at icosahedral 3-fold symmetry axes. These trimeric protrusions form isolated units that do not interact with each other. Broad depressions are observed near icosahedral 5-fold symmetry axes.

Structure of the HEV-CP. HEV-CP has a rather extended structure that can be divided into 3 linear domains: the S domain (118–313), the P1 domain (314–453), and the P2 domain (454–end) (Fig. 2*A* and *B*). Although HEV-CP contains 3 domains like the calicivirus coat protein, the organization of the 3 domains and their structural details are different, as described below.

The HEV-CP S domain adopts the typical jelly-roll β-barrel fold (14) that is widely conserved among many small RNA viruses. The signature 8 anti-parallel β strands are organized into BIDG and CHEF sheets, with 2 helices, α1 and α4, frequently found between strands C/D and E/F, respectively. A structural homolog search using the Dali server (15) showed that the S domain is best aligned with the jelly-roll domain of the coat protein of the carnation mottle virus (Z score = 17.5 with 2.0 being significant) (16), tomato bushy stunt virus (TBSV) (17), and other tombusviruses and sobemoviruses, all small T = 3 plant viruses with (+)RNA genomes. Animal viruses were ranked lower. Seneca Valley virus, an oncolytic picornavirus, had the highest ranking among all animal viruses (Z score = 14.7) (18). The capsid protein of Norwalk virus (19), a calicivirus that HEV was originally thought to be related to, exhibits even bigger differences in alignment than do picornaviruses (Z score = 11.8).

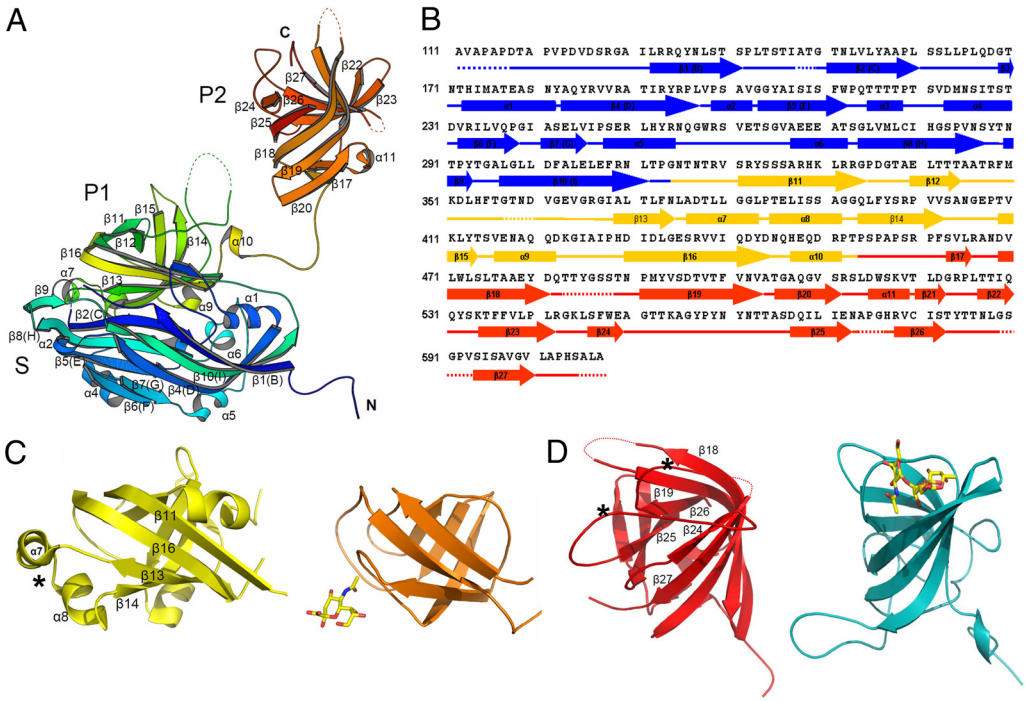

Fig. 2. The structure of HEV-CP. (*A*) One HEV-CP molecule. The molecule is rainbow colored with the N terminus in blue and the C terminus in red. (*B*) Secondary structure assignment. α-helices are shown by tubes, β-strands by arrows, loops by thick lines, and disordered regions by dotted lines. Regions from the S, P1, and P2 domains are colored in blue, yellow, and red, respectively, as in Fig. 1. The conventional naming scheme for the 8 β-strands (BIDG and CHEF) from the jelly-roll β-barrel is shown in parentheses. (*C*) P1 domain and endosialidase β-barrel domain. The sialic acid molecule is shown by sticks. Compared to (*A*), the P1 domain is rotated along the vertical axis by approximately 90° [viewed from left in *A*] to get a better view of the sialic acid binding site. (*D*) P2 domain and norovirus receptor binding domain. The blood-group polysaccharide is shown by sticks. The P2 domain has also been rotated compared to (*A*) to get a better view of the polysaccharide binding site. Secondary structural elements in P1 and P2 for putative sugar binding are highlighted by asterisks.

The HEV-CP P1 domain has the appearance of a squashed β-barrel consisting of 6 anti-parallel β-strands. The 2 ends of the barrel are flanked by several α-helices (Fig. 2 *A* and *C*). One side of the barrel extensively interacts with the S domain through βB, βC, and loops CD, EF, and GH. The structure of the HEV-CP P1 domain is related to that of the P2 domain of the calicivirus coat protein (Z score = 2.8) (20). Other top structural homologs are the human UPF1 human helicase core (Z score = 5.1, 1st) (21), the β-barrel domain of endo-alpha-sialidase (22), the tRNA-binding domain of the translation elongation factor Tu (23), and the receptor-binding domain of the avian reovirus fiber σC (24). Of these top homologs, the endosialidase of bacteriophage K1F is most interesting due to its ability to bind sialic acid molecules, which are widely distributed in animal tissues and bacteria (22) (Fig. 2*C*). By structural superposition, the potential sialic acid binding site of HEV-CP P1 is mapped to a helix-turn-helix motif (376–391) located at one end of the β-barrel (Fig. 2*C*).

The HEV-CP P2 domain forms the dimeric spike on the surface of the capsid (Figs. 1 and 2*D*). The overall fold of P2 is similar to that of P1, except for a large insertion (from 504–533) between β20 and β22 from the central β-barrel. This 30-amino acid insertion, comprised of 3 β strands and 1 α helix, mediates the interaction between the surface spike and the nearby 3-fold protrusion, thus helping support the spike. On top of the surface spike are 3 highly exposed large loop insertions (482–490, 550–566, and 583–593) that may play an important role in antigenicity determination. Superimposition of P2 and P1 by Dali server yields a Z score of 2.1. P2 is also homologous to the receptor binding domain of norovirus in the *Caliciviridae* family (Z score = 2.0) (25). Noroviruses use blood-group trisaccharides as cell receptors. Superimposing the 2 structures brought the trisaccharide to the top of the HEV surface spike between loops 550–566 and 583–593. Interestingly, this potential sugar binding site in P2 is structurally equivalent to the potential sialic acid-binding site in the P1 domain when the 2 domain structures are aligned together.

The P1 and P2 domains are connected by a long linker and do not interact directly. Structural flexibility in the linker is important because it allows the P2 domain to dimerize properly in the different types of dimers encountered in the native T = 3 capsid (see discussion below). Flexible hinge regions are also identified in the coat proteins of many T = 3 viruses, including TBSV and calicivirus, where dimers appear to be the assembly unit. Close inspection of the HEV linker sequence ^{445}NQHEQDRPTPSPAPSRPF^{462} indicates that it is rich in proline and thus a poor substrate for proteases in general. This explains our observation that protease treatment of the purified protein (with trypsin and chymotrypsin at up to a 1:1 mass ratio for 15 min at 20 °C) did not result in significant degradation. Evolutionary pressure brought about by the enteric transmission route of HEV may have selected for a capsid protein with sequence and structural features that make the virus highly resistant to proteases.

Residues 118–131 at the N terminus of the HEV-CP form the N-terminal arm (Fig. 2 *A* and *B*). The arm makes a sharp turn at the beginning of βB, forming an extended loop that interacts with a 2-fold related and then a 3-fold related molecule nearby. In TBSV, the N-terminal R segment has an important role in regulating particle size (14). It wedges between 2 capsid protein subunits, thereby creating flat dimers on icosahedral 2-fold axes. The N-terminal arm of HEV-CP is away from the 2-fold dimer interface and adopts a different conformation compared to the R segment of TBSV.

Structure of the T = 1 Subviral Particle. The 3 domains of HEV-CP play different roles in organizing the icosahedral capsid. The S domain, which adopts the jelly-roll, β-barrel fold, forms a continuous capsid shell with pointed vertices and flat facets (Figs. 1*C* and 3). There are no obvious holes or cavities, suggesting that the S domain alone may be capable of self assembly. The P1 domain interacts near 3-folds, forming isolated trimeric protrusions (Figs. 1*D* and 3). Thus, the P1 domain stabilizes only the trimeric interactions. The P2 domain forms dimeric spikes that stabilize capsid protein interactions across the 2-folds. The critical role of the S domain in capsid assembly is evident, as P1 and P2 cannot completely close the particle (Figs. 1*E* and 3). Approximately 1,400-Å^2 total surface area is buried between adjacent molecules near the 5-fold (S, ~1,400 Å^2; P1, none; P2, none), approximately 3,000 Å^2 around the 3-fold (S, ~2,000 Å^2; P1, ~1,000 Å^2; P2, none), and approximately 5,300 Å^2 around the 2-fold (S, ~2,500 Å^2; P1,

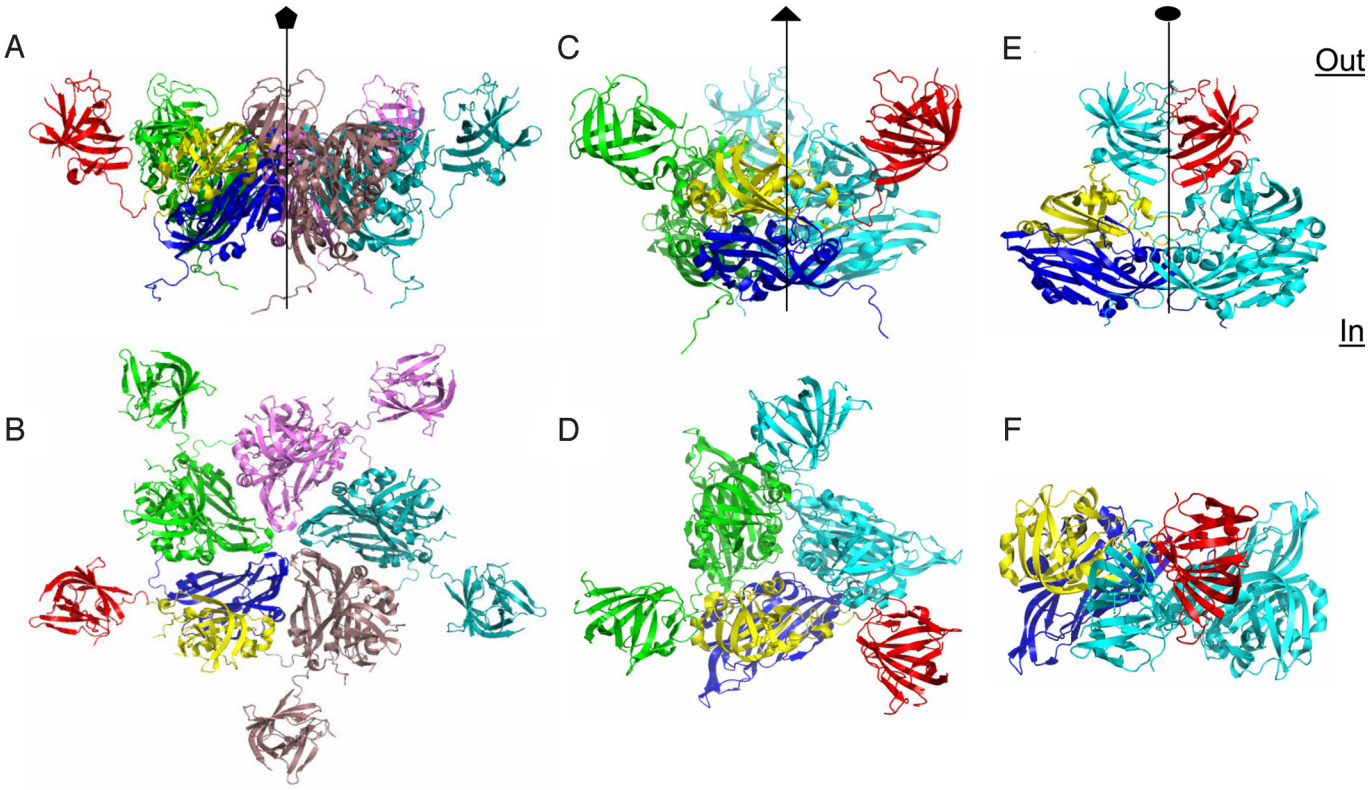

Fig. 3. HEV VLP capsomeres. (*A* and *B*) HEV-CP pentamer. A reference molecule is colored according to domain organization with S in red, P1 in yellow, and P2 in red. Other molecules are shown with a single color only. (*C* and *D*) HEV-CP trimer. (*E* and *F*) HEV-CP dimer. (*A, C,* and *E*) are viewed from the side, whereas (*B, D,* and *F*) are viewed along the symmetry axes from the outside of the VLP.

~200 Å^2; P2, ~2,600 Å^2). Based on the buried surface areas, the 5-fold interaction is the weakest, and the 2-fold interaction is most stable.

The P2 domain alone is capable of dimerization (26). Dimerization of the P2 domain is mediated by an extended loop (550–566) and 3 β-strands from the central β-barrel (β18, β24, and β27). The 4 structural elements provide a flat interface that is largely hydrophobic in nature. Previous mutagenesis identified a cluster of 6 hydrophobic residues critical for dimeric interactions: A597, V598, A599, L601, and A602 (26). Another study found that the deletion of residues 585–610 led to reduced oligomerization and aberrant folding of the protein (27). The VLP structure shows that residues 594–600 form one of the β-strands (β27) at the dimer interface. Amino acid substitutions within this β-strand are likely to affect either the folding or properties of the interface, thus resulting in the disruption of the tight packing between the 2 β-sheets.

The VLP structure allows us to evaluate the potential physiological relevance of the 3 potential N-glycosylation sites (5). N137 is partially hidden near the inner surface of the capsid shell and N310 is completely buried, suggesting that glycosylation at either site is unlikely. N562 is exposed to solvent at the very top of the surface spike and could potentially be subjected to glycosylation in the ER. The inner surface of the capsid shell is covered with a large number of basic amino acid side chains (R128, R133, R186, R189, R193, and R195, 6 from each subunit), remarkably different from dsRNA viruses in which a large number of *negatively* charged residues on the inner surface are used to facilitate the movement of dsRNA genome during particle-associated transcription [for an example, see (28)]. These arginine side chains from HEV-CP presumably help to neutralize the negative charges of the genomic RNA. Around the 5-fold axes is a ring of 5 tyrosine residues (Y288) that are hydrogen bonded to 5 serine residues (S200), which are also positioned around the 5-fold axes, but closer to the particle interior. We speculate that the dissociation of the VLP at alkaline pH may have been caused by de-protonation of the tyrosine side chain, resulting in the destabilization of the 5-fold interaction.

Given that a potential sugar binding site is found in both the HEV-CP P1 and P2 domains, it is important to determine which of the 2 domains functions in cell receptor binding. An earlier study showed that the ORF2$^{368-606}$ protein was able to bind and penetrate different cell lines susceptible to HEV and to inhibit HEV infection (8). The ORF2$^{368-606}$ protein, however, contains both potential receptor binding sites. The potential sugar binding sequence from the P1 domain, ^{376}ADTLLGGLPTELISSA^{391}, is strictly conserved among all 4 HEV genotypes, suggesting that it has an important functional role in cell receptor binding. In contrast, loops 550–566 and 583–593, the potential sugar binding site of the P2 domain, contain 3 and 4 hypervariable amino acid sites, respectively, indicating that these regions are instead likely to mediate antibody recognition and immune escape. Moreover, the putative sugar binding motif in the P1 domain (376–391) forms a hidden pocket at the interface between 2 HEV-CP molecules around the 3-fold. Therefore, receptor binding to P1 may potentially lead to the destabilization of the HEV-CP trimer, resulting in conformational changes that eventually lead to membrane penetration and genome release into the infected cell. Further mutagenesis studies targeting these 2 potential sugar-binding sites will determine which domain functions in host-cell binding and virus internalization.

A Native T = 3 Capsid Model. Based on the T = 1 HEV VLP structure, we have modeled a T = 3 capsid by aligning the S domain of HEV with that of the 3 quasi-equivalent capsid protein molecules, A–C from TBSV (17) (Fig. 4*A*). TBSV is one of the closest structural homologs of HEV-CP according to Dali. Previous studies on small plant RNA viruses, insect nodaviruses, and caliciviruses all indicate that the assembly of T = 3 viral capsid requires 2 different

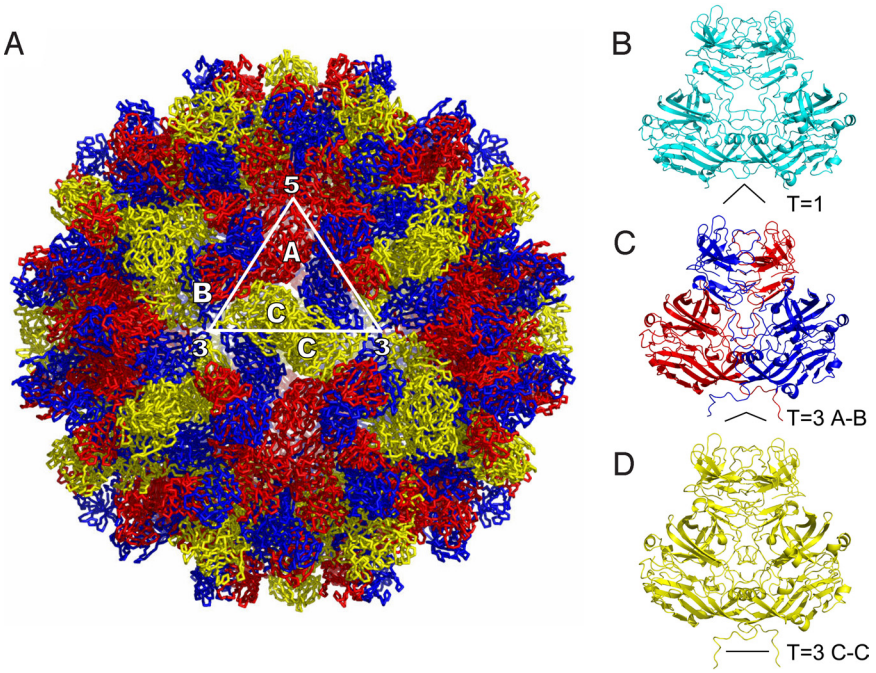

Fig. 4. T = 3 HEV capsid model. (*A*) The overall structure. The 3 quasi-equivalent CP molecules A–C are colored blue, red, and yellow, respectively. One asymmetric unit is highlighted along with icosahedral symmetry axes. (*B*) T = 1 VLP dimer. (*C*) T = 3 A-B dimer. (*D*) T = 3 C-C dimer. All 3 dimers are viewed from the side of their 2-fold symmetry axes. The line drawings at the bottom of the CP dimers illustrate the different hinge angles observed in different dimers.

forms of dimers: one that is flat, sitting on 2-fold axes, and another that has an inwardly bent conformation, located on quasi 2-folds (14). Therefore, it is likely that the assembly of a T = 3 HEV capsid follows the same principle. By superimposing the S domains, the HEV-CP P1 domain is also brought close to quasi 3-fold axes, forming trimeric protrusions similar to those seen in T = 1 VLP. The P2 domain, however, needs manual adjustments to make close dimer contacts. Such variations in P2 orientation can easily be accommodated by the flexible hinge linker between P1 and P2.

The T = 3 HEV capsid has an inner diameter of approximately 220 Å and an outer diameter of approximately 370 Å, consistent with previous immunoelectron microscopy observation (1). Large depressions are found on 5-fold as well as quasi 6-fold symmetry axes. There are a total of 90 dimeric surface spikes (2 types, C-C and A-B) and 60 trimeric protrusions (1 type, A-B-C). Comparison of the C-C and A-B dimers to the T = 1 VLP dimer reveals that the C-C and A-B dimers are flatter, with the C-C dimer being the flattest among the three (Fig. 4 *B–D*). The reason that recombinant T = 3 HEV particles have not been observed is likely because the flat dimers are energetically unfavorable in the absence of genomic RNA and/or the rest of the N-terminal arm. A role for the N-terminal arm is supported by the fact that the N-terminal peptide 118–131 needs to adopt a different configuration in the T = 3 particle: in our T = 3 model the peptide points toward the interior of the particle and does not make contacts with neighboring molecules. As the interactions around the 5-fold and quasi 6-fold are weak, the assembly of HEV T = 3 capsid likely proceeds through the formation of trimers of dimers, which then oligomerize near the 5-fold and quasi 6-fold symmetry axes to complete the assembly process. A similar scheme has been proposed for tombus viruses (e.g., turnip crinkle virus) (29), but calicivirus appears to follow an alternative pathway involving pentamers of dimers (19).

The capsid structure of HEV differs from that of calicivirus in several major respects. First, the 3 domains S, P1, and P2 are arranged in a linear sequence in HEV-CP. In calicivirus, the P2 domain is a large insertion in the P1 domain. Second, the structural folds of the P1 domains in the 2 viruses are unrelated. Third, the P1 domain of HEV-CP forms trimeric protrusions whereas the P1 domain of calicivirus capsid protein interacts near the icosahedral 2-folds. Fourth, the HEV-CP S domain is more closely related to plant T = 3 viruses, whereas the calicivirus capsid protein is more similar to picornaviruses. For example, a Dali search found that the best structural homolog of the Norwalk virus capsid protein S domain (excluding other members of the *Caliciviridae* family) was human rhinovirus serotype 2 (30) (Z score = 15.5), closely followed by coxsackievirus (31), cricket paralysis virus (32), and other picorna or picorna-like viruses. Plant T = 3 viruses are ranked much lower with the first being sesbania mosaic virus (Z score = 12.5) (33), which is a sobemovirus. Last, HEV codes for a methyltransferase so its genome should have a 5′-cap like plant T = 3 viruses (1). Caliciviruses, however, use protein-primed RNA synthesis and have a VPg-linked genome like picornaviruses. These differences suggest that HEV and caliciviruses may have evolved through 2 separate lineages that are each related to small plant (+)RNA viruses and picornaviruses, respectively. It would be interesting whether further structural analysis of common nonstructural proteins from HEV, TBSV, and calicivirus would also lead to the same conclusion once these data become available.

Our HEV VLP structure provides an important framework to better understand and model the molecular biology of the hepatitis E virus. Our mechanistic models for HEV capsid assembly, cell receptor binding, and antibody neutralization can be easily tested by site-directed mutagenesis. Because recombinant HEV-CP can readily assemble into particles, this system is also highly amenable to in vitro studies of virus assembly. This HEV VLP structure will help researchers develop significant applications such as vaccines and antiviral medications to better control HEV infection in humans and animals.

Materials and Methods

Cloning, Expression, and Purification. The genetic sequence of the HEV (China T1 Isolate, genotype 4) ORF2 encoding residues 112–608 of the capsid protein (abbreviated as HEV-CP) was cloned into pFastBac1 (Invitrogen). A 6-residue histidine tag was added to the C terminus. Infected Sf21 insect cells were harvested 48 h post-infection. Cell pellets were sonicated in lysis buffer containing 50 mM Tris-HCl, pH 7.5, 300 mM NaCl, and 5 mM imidazole. Recombinant HEV-CP was purified by Ni-NTA affinity (Qiagen), heparin affinity, Superdex-200 gel filtration, and Q anion exchange columns (GE). Purified protein was concentrated and stored at 4 °C in buffer containing 50 mM Tris-HCl, pH 7.5, 200 mM NaCl, and 5 mM β-mercaptoethanol.

Cryo-EM and Image Reconstruction. The VLPs for cryoEM were obtained by expressing the HEV-ORF2$^{112-660}$ protein (HE-JF4 isolate, genotype 4) in Tn5 insect cells as previously described (10). Gradient-purified VLP samples were diluted to about 2.5 mg/mL with PBS (pH 7.2) and plunge-frozen on Quantifoil grids. Images were collected on film using a JEM2011 electron microscope at 200 kV and 50,000× magnification. Defocus values of approximately 1.0 and 3.0 μm were used. Micrographs were digitized at 1.27 Å/pixel. Data processing was carried out using the IMIRS package (34). Particle orientations were first determined by Fourier common line and subsequently refined by projection-matching. Image defocus values were estimated using the Ctfit from the EMAN package (35). A total of 1,051 particle images were used to reconstruct the final map. The effective resolution was estimated to be 14-Å resolution (FSCC = 0.5).

Crystallization, Data Collection, and Data Processing. Crystals of HEV-CP were obtained using vapor diffusion. The hanging drop consisted of equal volumes of protein (10 mg/mL) and well solution (0.1 M sodium cacodylate, pH 6.5, and 1.5 M lithium sulfate). Hexagonal-shaped crystals appeared after 2 weeks of incubation at 20 °C and grew to full size (0.4 × 0.4 × 0.1 mm^3) after 2 months. Microseeding and the addition of the detergent n-tetradecyl-b-D-maltoside accelerated crystal growth and improved crystal quality. Diffraction images were collected (oscillation angle = 0.5°) from frozen crystals presoaked in cryoprotectant made of well solution with 25% glycerol. Diffraction data were processed using the program HKL2000 (36). HEV-CP crystals had the space group of $P6_3$ with a = 241.1 Å and c = 519.9 Å. Our final data set was obtained from 2 crystals indexed in the same handedness (Table S1).

Structure Determination. Self-rotation function using the program GLRF (37) showed peaks consistent with icosahedral symmetry. Packing consideration further suggested that there were 2 VLPs in each unit cell, with 1/3 of the particle in each asymmetric unit. The icosahedral 3-fold symmetry axis of the VLP was aligned with the crystallographic 6_3 symmetry axis. Of the 2 possible crystal packing arrangements with VLP centered at (0, 0, 0) or (1/3, 2/3, 0), the former gave better crystal contact considering that HEV VLP has a diameter of only approximately 270 Å. Of the 6 parameters describing VLP position and orientation, the only unknown was the rotation angle of the VLP around the crystallographic 6_3 symmetry axis. Using diffraction data from 4.5- to 3.5-Å resolution, self-rotation function determined that 1 icosahedral 5-fold axis was at ϕ = 49.20 °, ψ = 37.38 °.

A 14-Å cryoEM reconstruction, flattened to a binary map with densities corresponding to the capsid shell and solvent regions set to 1 and 0, respectively, was used for initial phasing. EM map magnification and the contour level for map flattening were optimized by maximizing correlation between the phasing model and the diffraction data. Using 20-fold NCS averaging, phases were gradually (one reciprocal lattice interval per step) extended to 3.5 Å using the programs RAVE (38) and CCP4 (39), with 10 cycles of averaging at each phase extension step. To optimize NCS matrices, the same phase extension procedure was repeated by manually adjusting the icosahedral 5-fold axis with ϕ varying from 48.90° to 49.50° at 0.05° intervals. The observation of right-handed α-helices confirmed the correct handedness of our map. The final map was sharpened using B = −150 Å^2 (Fig. S2). Atomic models were built in O (40) and subsequent refinements were performed using CNS (41) with 20-fold NCS constraints. The Ramachandran plot calculated for the final model showed 82% of non-Gly residues in the most-favored regions with none in disallowed regions based on main-chain dihedral angles.

Ribbon diagrams and Cα traces were prepared using Molscript (42), PyMOL (W.L. Delano, http://www.pymol.org), and Chimera (43).

ACKNOWLEDGMENTS. We thank B.V.V. Prasad, J. Pan, R.R. Reed, and C. Ke for valuable discussions, and the staff at CHESS and APS for assistance with data collection. This work was supported by the National Institutes of Health (Y.J.T.), the National Natural Scientific Foundation of China (J.Z.), the Major State Basic Research Development Program of China (C.Y.), the Major State Science and Technology Project of China (C.Y.), the Welch Foundation (Y.J.T.), the Hamill Foundation (Y.J.T.), and by the Kresge Science Initiative Endowment Fund at Rice University.

1. Emerson UE, Purcell RH (2007) in *Fields Virology*, eds Knipe,D, Howley P (Lippincott Williams & Wilkins, Philadelphia), Vol 1, pp 3047–3058.
2. Chandra V, Taneja S, Kalia M, Jameel S (2008) Molecular biology and pathogenesis of hepatitis E virus. *J Biosci* 33:451–464.
3. Tam AW, et al. (1991) Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral genome. *Virology* 185:120–131.
4. Surjit M, Jameel S, Lal SK (2004) The ORF2 protein of hepatitis E virus binds the 5′ region of viral RNA. *J Virol* 78:320–328.
5. Zafrullah M, Ozdener MH, Kumar R, Panda SK, Jameel S (1999) Mutational analysis of glycosylation, membrane translocation, and cell surface expression of the hepatitis E virus ORF2 protein. *J Virol* 73:4074–4082.
6. Graff J, et al. (2005) In vitro and in vivo mutational analysis of the 3′-terminal regions of hepatitis e virus genomes and replicons. *J Virol* 79:1017–1026.
7. Torresi J, Li F, Locarnini SA, Anderson DA (1999) Only the non-glycosylated fraction of hepatitis E virus capsid (open reading frame 2) protein is stable in mammalian cells. *J Gen Virol* 80:1185–1188.
8. He S, et al. (2008) Putative receptor-binding sites of hepatitis E virus. *J Gen Virol* 89:245–249.
9. Zhang Y, McAtee P, Yarbough PO, Tam AW, Fuerst T (1997) Expression, characterization, and immunoreactivities of a soluble hepatitis E virus putative capsid protein species expressed in insect cells. *Clin Diagn Lab Immunol* 4:423–428.
10. Xing L, et al. (1999) Recombinant hepatitis E capsid protein self-assembles into a dual-domain T = 1 particle presenting native virus epitopes. *Virology* 265:35–45.
11. Khudyakov YE, et al. (1999) Antigenic domains of the open reading frame 2-encoded protein of hepatitis E virus. *J Clin Microbiol* 37:2863–2871.
12. Meng J, et al. (2001) Identification and characterization of the neutralization epitope(s) of the hepatitis E virus. *Virology* 288:203–211.
13. Shrestha MP, et al. (2007) Safety and efficacy of a recombinant hepatitis E vaccine. *N Engl J Med* 356:895–903.
14. Harrison SC (2007) in *Fields Virology*, eds Knipe D, Howley P (Lippincott Williams & Wilkins, Philadelphia), Vol 1, pp 59–98.
15. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v 3. *Bioinformatics* 24:2780–2781.
16. Morgunova E, et al. (1994) The atomic structure of Carnation Mottle Virus capsid protein. *FEBS Lett* 338:267–271.
17. Olson AJ, Bricogne G, Harrison SC (1983) Structure of tomato busy stunt virus IV: The virus particle at 29 A resolution. *J Mol Biol* 171:61–93.
18. Venkataraman S, *et al.* (2008) Structure of Seneca Valley Virus-001: An oncolytic picornavirus representing a new genus. *Structure* 16:1555–1561.
19. Prasad BV, *et al.* (1999) X-ray crystallographic structure of the Norwalk virus capsid. *Science* 286:287–290.
20. Chen R, Neill JD, Estes MK, Prasad BV (2006) X-ray structure of a native calicivirus: structural insights into antigenic diversity and host specificity. *Proc Natl Acad Sci USA* 103:8048–8053.
21. Cheng Z, Muhlrad D, Lim MK, Parker R, Song H (2007) Structural and functional insights into the human Upf1 helicase core. *EMBO J* 26:253–264.
22. Stummeyer K, Dickmanns A, Muhlenhoff M, Gerardy-Schahn R, Ficner R (2005) Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. *Nat Struct Mol Biol* 12:90–96.
23. Andersen GR, Thirup S, Spremulli LL, Nyborg J (2000) High resolution crystal structure of bovine mitochondrial EF-Tu in complex with GDP. *J Mol Biol* 297:421–436.
24. Guardado-Calvo P, et al. (2008) Crystal structure of the avian reovirus inner capsid protein sigmaA. *J Virol* 82:11208–11216.
25. Cao S, et al. (2007) Structural basis for the recognition of blood group trisaccharides by norovirus. *J Virol* 81:5949–5957.
26. Li TC, et al. (2005) Essential elements of the capsid protein for self-assembly into empty virus-like particles of hepatitis E virus. *J Virol* 79:12999–13006.
27. Xiaofang L, Zafrullah M, Ahmad F, Jameel S (2001) A C-Terminal Hydrophobic Region is Required for Homo-Oligomerization of the Hepatitis E Virus Capsid (ORF2) Protein. *J Biomed Biotechnol* 1:122–128.
28. Pan J, et al. (2009) Atomic structure reveals the unique capsid organization of a dsRNA virus. *Proc Natl Acad Sci USA* 106:4225–4230.
29. Sorger PK, Stockley PG, Harrison SC (1986) Structure and assembly of turnip crinkle virus II Mechanism of reassembly in vitro. *J Mol Biol* 191:639–658.
30. Verdaguer N, Blaas D, Fita I (2000) Structure of human rhinovirus serotype 2 (HRV2). *J Mol Biol* 300:1179–1194.
31. Muckelbauer JK, et al. (1995) The structure of coxsackievirus B3 at 35 A resolution. *Structure* 3:653–667.
32. Tate J, *et al.* (1999) The crystal structure of cricket paralysis virus: the first view of a new virus family. *Nat Struct Biol* 6:765–774.
33. Bhuvaneshwari M, *et al.* (1995) Structure of sesbania mosaic virus at 3 A resolution. *Structure* 3:1021–1030.
34. Liu H, *et al.* (2008) Symmetry-adapted spherical harmonics method for high-resolution 3D single-particle reconstructions. *J Struct Biol* 161:64–73.
35. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128:82–97.
36. Otwinowski Z, Minor W (1997) in *Methods in Enzymology*, eds Carter CW, Sweet RM (Academic, New York), Vol 276, pp 307–326.
37. Tong L, Rossmann MG (1997) Rotation function calculations with GLRF program. *Methods Enzymol* 276:594–611.
38. Jones TA (1992) in *Molecular Replacement*, eds Dodson EJ, Gover S, Wolf W (SERC Daresbury Laboratory, Warrington), pp 91–105.
39. CCP4 (Collaborative Computational Project Number 4) (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
40. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110–119.
41. Brunger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
42. Kraulis PJ (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
43. Pettersen EF, et al. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.